# PDS4 Data Standards

## Steve Hughes

MC Face-to-Face
Tucson, Arizona
April 10-11, 2014

# Topics

- Status of the PDS4 Information Model (IM) with respect to PDS4 Build 4b

  - Standards Documents Statuses

- Data Design Working Group and Task Statuses

- Other work

  - External requests for extracts from the IM
  - Papers and presentations at conferences

# Summary of Progress
## (Since Release of V1.1.0.1)

- Released IM V1.2.0.0 for Build 4b for Model Integration and Testing on 3/17/14

- Released IM V1.2.0.1 for System Integration and Testing on 3/31/14

- Priority given to requirements for LADEE, MAVEN, InSight and Osiris-Rex
  - Insight/SEED Data Format
  - Osiris-Rex Geometry

# Release Process for the IM

- Update IM with approved change requests
  - Change request is prototyped using a design tool to determine viability.
  - Technical assessment is written.
  - If approved by CCB, the change is made permanent in a new version of the IM.
- Generate documents from the IM
- Perform difference and regression tests
- Release to PDS4 "develop" directories for Unit Testing
  - Fix bugs found during Unit Testing
  - Nodes are requested to participate
- Release to PDS4 "released" directories for Model Integration and Testing (MI&T)
  - Fix bugs found during MI&T
- Release for System Integration and Testing

# CCB approved changes

- CCB-20 - Correct Minor Errors in Enumerated Values
- CCB-22 – Adopt Rules for Case of Enumerated Values - (Part 2)
  - Part 1 - Completed for V1.1.0.0
  - Part 3 – Steward_Id - no changes necessary
- CCB-29 - Deprecate Display_2D_Image class
  - Add Display Dictionary
- CCB-30 - Create Local_Internal_Reference class
- CCB-41 - (Part 2) – Add schematon rules for fields and groups both not being zero
- CCB-42 - CCB-31 incorrectly implemented
- CCB-45 - Document use of Science_Facets class
- CCB-46 – Add Product_Native for InSight/SEIS SEED data format
            (Policy issue resolution still required)

# Documents[1]

**Standards Documents (released for Build 4b)**

- Information Model Specification – Version 1.2.0.1
- XML Schemas – Version 1.2.0.1
- Data Dictionary – Version 1.2.0.1
- PDS4 Example Products - Version 1.2.0.1
- Standards Reference - Version 1.2.0

**Support Documents (planned reviews and updates)**

- DD Tutorial
- Glossary
- Concepts Document
- Data Provider's Handbook
- PAG (Proposers Archive Guide)

- APG (Archive Preparation Guide) – Replaced by DPH

6

[1]*Posted to* *http://pds.nasa.gov/pds4*

# DDWG Status and Tasks

- The DDWG now meets for one hour[1] teleconferences on alternate weeks
- Agenda
  - Discipline extensions to the IM
  - Advise on CCB change request issues
  - Maintain IM common core
- Discipline Team Tasks
  - Geometry (focus: flyby and orbital missions)
  - Cartography
  - NSSDC/PDS Interface
  - Document Improvement
  - *Value Meaning Improvement*

[1]*Has been know to exceed the allocated time*

# Task - Geometry

**Team lead and members:** <u>E. Guinness, M. Gordon</u>, A. Raugh, T. Farnham, C. Isbell, S. McLaughlin, B. Semenov , C. Acton, E. Rye, S. Hughes

**Short Description:** Capture geometry requirements from across the disciplines, obtain a consensus model, and write a geometry dictionary.  Currently focusing on classes for flyby and orbital missions.

**Goals:** Cross-discipline Geometry Model and Dictionary

**Schedule (Major milestones):**
  Mar – Start gathering requirements
  Apr – Draft Dictionary for Osiris-Rex
  Jul – Release final version

# Task - Cartography

**Team lead and members:** <u>E. Rye</u>, C. Isbell

**Short Description:** Using existing and accepted cartography standards  extend/develop a PDS4 cartography model that can be used across the PDS discipline nodes and also extended as needed for specific purposes.

**Goals:** Develop a cartography model suitable for use across the PDS discipline nodes.

**Schedule (Major milestones):**
  Apr – Continue Development
  May – Test ingest Cartography Model into PDS4
  Sep – Release for general use in next Build.

# Task - Document Improvement - 1

**Team members:** D. Simpson, M. Gordon, R. Joyner

**Short Description:** Address issues associated with the PDS4 Data Standards documents.

**Goals:** Consistent data standard documents and information model.

**Schedule (continuing work):**
- Support Documents – Planned Reviews and Updates
    - DD Tutorial
    - Glossary and List of Acronyms and Abbreviations
    - Concepts Document
    - Data Provider's Handbook
    - PAG

# Task- Document Improvement - 2

**Schedule (continuing work):**

- Receive comments on SR from IPDA

- Follow through with revisions of enumerated values and related issues in the Standards Reference

# Task - NSSDC/PDS Interface

**Team lead and members:**   S. McLaughlin, E. Bell, P. McCaslin, J. Kodis, S. Hughes,  S. Hardman, E. Guinness, L. Huber, C. Isbell, T. King, A. Raugh

**Short Description:**   Develop an interface for transferring PDS4 data to the NSSDC deep archive

**Goals:**

- Automate deliveries and eliminate multiple ingests of a given product
- Provide the ability to return basic products, collections, and bundles from the deep archive and report on those entities
- Perform data integrity checks

# Task - NSSDC/PDS Interface

**Schedule (Major milestones):**

***Past 6 months***

- Reached agreements on NSSDC services, PDS and NSSDC expectations and responsibilities
- Launched a test Registry Service populated with a Phoenix bundle
- Designed a preliminary PDS4 delivery package
- Began identifying changes to the NSSDC ingest data model and applications
- Outlined a delivery process that frees the NSSDC of accessing the Registry Service for transporting data

***Next several months***

- Finalize the PDS4 delivery package and process
- Continue development of interface and begin testing

# Information Model Extracts

- Requests have been made from projects for extracts of the contents of the PDS4 Information Model in other machine languages.

  - JSON[1] – APPS (AMMOS-PDS Pipeline Service)
  - SKOS[2] and OWL[5]– Linked Open Data[3] Project (Bernd Ritschel – GFZ Potsdam)
  - RDF[4] – Various requesters

[1] JavaScript Object Notation (JSON) is a lightweight data-interchange format.
[2] Simple Knowledge Organization (SKOS) is a common data model for sharing and linking knowledge organization systems via the Web.
[3] Linked Open Data (LOD) is a way of publishing structured data that allows metadata to be connected and enriched, so that different representations of the same content can be found, and links made between related resources.
[4] Resource Description Framework (RDF) is a family of World Wide Web Consortium (W3C) specifications used as a general method for conceptual description or modeling of information that is implemented in web resources, using a variety of syntax notations and data serialization formats.
[5] The Web Ontology Language (OWL) is a family of knowledge representation languages or ontology languages for authoring ontologies or knowledge bases.

# Papers and Presentations - 1

- Presented the paper, *PDS4: A Model-Driven Planetary Science Data Architecture for Long-Term Preservation,* at the Workshop on Long Term Preservation for Big Scientific Data

  - Associated with the 30th IEEE International Conference on Data Engineering.

  - Most of the those present were particle physicists associated with CERN and other colliders. They were very interested in information model development since currently they have none and now need to preserve their data.

  - Requested to submit an abstract to the 10th IEEE E-Science Conference by one of the e-science conference program chairs.

# Papers and Presentations - 2

- Prepared *Information Model Translation to Support a Wider Science Community* for presentation at the European Geosciences Union General Assembly, 2014.
  - Presentation will be made by co-author B. Ritschel.
  - Component of the Linked Open Data project.
    - The PD4 information model contents are written to RDF-based data serialization formats, SKOS and OWL.
    - Co-author on another presentation.

- Participated on a panel for *Developing and Implementing Institutional Policies on Research Data*
  - Associated with the Research Data Access and Preservation Summit 2014 (RDAP14).
  - Presented *Policy Recommendations for Institutions to Serve as Trustworthy Stewards of Research Data.*
  - Member of the Primary Trustworthy Digital Repository Authorisation Body (PTAB)

# Questions and Answers

# Backup

# Task - NSSDC/PDS Interface

**The Team**
NSSDC: S. McLaughlin, E. Bell, P. McCaslin, J. Kodis, M. Martin
PDS: S. Hughes,   S. Hardman, E. Guinness, L. Huber, C. Isbell, T. King, A. Raugh

**Short Description & Goals**
Develop an interface for transferring PDS4 data to the NSSDC deep archive
- Automate deliveries and eliminate multiple ingests of a given product
- Provide the ability to return basic products, collections, and bundles and report on those entities
- Perform data integrity checks

**Schedule/Major Milestones**
*Past 6 months*
- Reached agreements on NSSDC services, PDS and NSSDC responsibilities
- Launched a test Registry Service populated with a Phoenix bundle
- Designed a preliminary PDS4 delivery package
- Began identifying changes to the NSSDC ingest data model and applications
- Outlined a delivery process that frees the NSSDC of accessing the Registry Service for transporting data

*Next 2 months*
- Finalize the PDS4 delivery package and process
- Continue development at the NSSDC and EN
- Test ingest of Cartography Model into PDS4 (an EN activity?)

*This summer & fall*
- Continue development and testing
- Release for general use in next Build

# Testing

- Oxygen is used to test the generated XML Schemas and Schematron files for validity and well-formedness.
- File Difference (Ultra-Edit Compare) is used to identify differences between current and prior versions of generated files after updates.
  - XML Schema, Schematron, Information Model Specification, ISO/IEC-11179 Data Dictionary files.
- Regression tests are performed against a suite of example products.
- Regression tests are performed against selected node bundles (Imaging, Atmospheres)

# PDS4 Documents
# and their Relationships
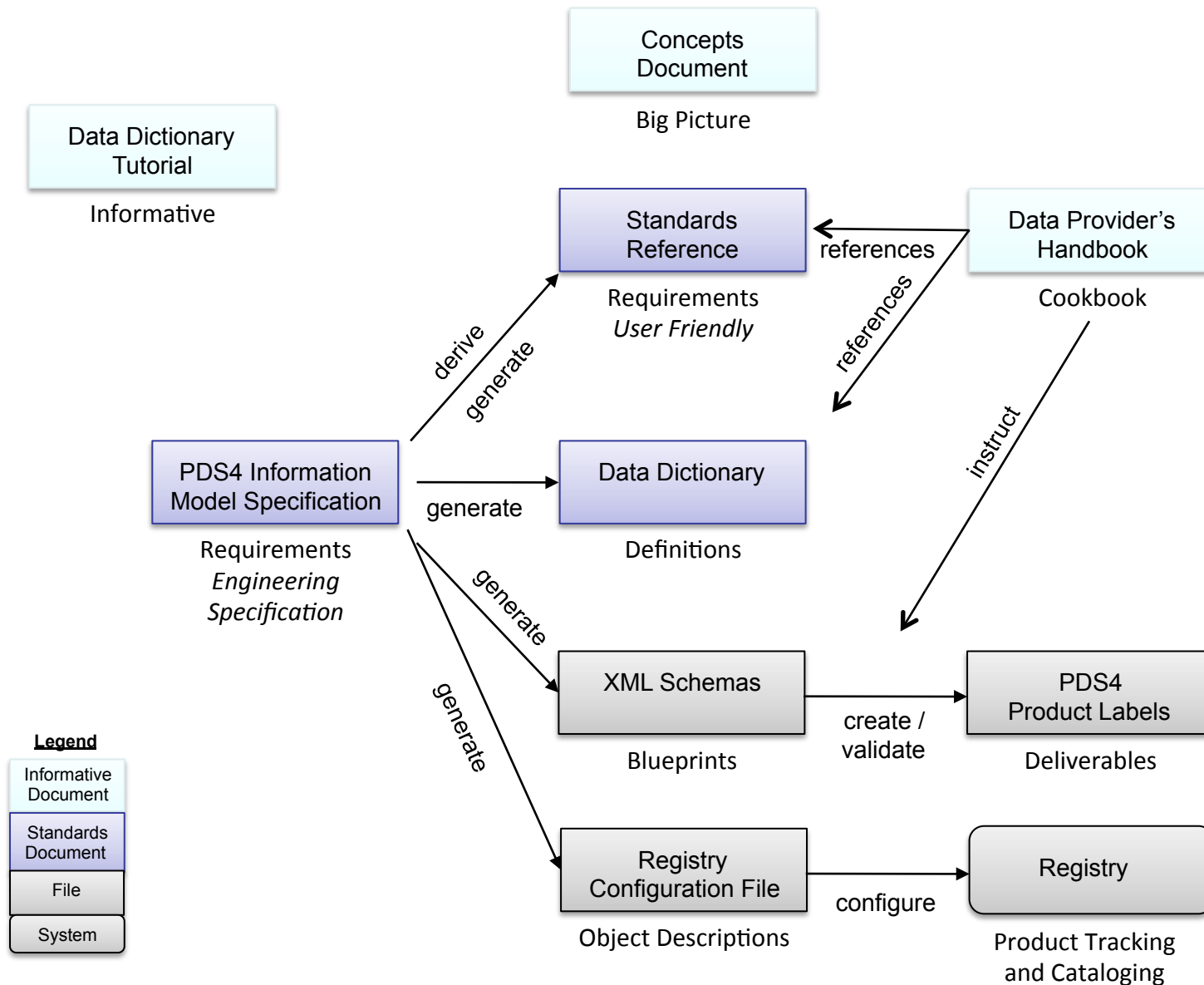
# Acknowledgements*

Ed Bell
Richard Chen
Dan Crichton
Amy Culver
Patty Garcia
Ed Grayzeck
Ed Guinness
Mitch Gordon
Sean Hardman
Lyle Huber
Steve Hughes
Chris Isbell
Steve Joy
Ronald Joyner

Debra Kazden
Todd King
John Kodis
Joe Mafi
Mike Martin
Thomas Morgan
Lynn Neakrase
Paul Ramirez
Anne Raugh
Shannon Rees
Mark Rose
Matias Roybal
Elizabeth Rye
Boris Semenov

Dick Simpson
Susie Slavney
Dillon White

Peter Allan
David Heather
Michel Gangloff
Santa Martinez
Thomas Roatsch
Alain Sarkissian

\* Anyone who sat through a DDWG 2-hour telecon or provided useful input.